用于探测深层语义的自然语言分解算法

刘廷镇
腾讯 IEG，研发效能部
firstsg@outlook.com

摘要

认知科学研究认为，文本在被逻辑处理之前，首先会形成一种下意识理解。而这种下意识理解的“概念域”基于我们所处的社会文化环境而产生。

文本传达的下意识情绪是表达中的重要组成部分，在文学写作、政治宣传、心理治疗等领域有着重要意义。然而目前文本倾向性分析研究主要着重于对文本进行逻辑理解，并不关注文本潜在表达的识别与构造。

本文提出了一种一体两面的文本分解与文本生成算法，利用预训练语言模型分析与构造文本的潜在表达。不同于基于神经网络迁移学习的工作，本文的方法不需要对语言模型进行微调，而是利用语言模型表征的信息构建语义空间，并在语义空间上传播标记信息，这使得算法避免了深度学习的昂贵计算开销与训练的不稳定性。与目前多维度情感分析工作相比，文本分解部分允许用户自定义不同的分解目标，基于半监督学习方法，仅仅需要进行少量标记即可训练将文本分解为目标情绪分量上组分的可靠模型。文本生成部分可以基于分解部分标注的模型生成具有特定情绪分量的文本。与目前文本风格迁移工作相比，文本的生成模型不需要使用神经编解码器，因此有更高的效率和更好的可解释性。

本文描述的算法是一套前瞻性的工具内核，其拥有可解释的推理过程，通过定义不同的分解目标，该方法可以被用于多维度情感分析、舆情监测等多种任务，也为大规模、自动化心理测量提供了一种可能。

关键词： 自然语言处理；文本生成；多维度文本分析；语言模型；半监督学习

Natural Language Decomposition Algorithm for Deep Semantic Analysis

Tingzhen Liu
Tencent IEG, Frontier Technology Department
firstsg@outlook.com

Abstract

Cognitive science research holds that a text will first form a kind of subconscious understanding before it is logically processed. This "conceptual domain" of subconscious understanding is based on our social and cultural environment.

The subconscious emotion conveyed by the text is an important part of expression, which is of great significance in the fields of literary writing, political propaganda, psychotherapy and so on. However, the current research on text propensity analysis mainly focuses on the logical understanding of the text, and does not pay attention to the identification and construction of the potential expression of the text.

In this paper, an algorithm of text decomposition and text generation is proposed, which uses the pre-training language model to analyze and construct the potential expression of text. Different from the work based on neural network transfer learning, the method in this paper does not need to fine-tune the language model, but uses the representation by the language model to build the semantic space. And propagates the tag information in the semantic space. This method avoids the expensive computational cost and the instability of training of deep learning. Compared with the current multi-dimensional emotion analysis work, the text decomposition part allows users to customize different decomposition goals. Based on the semi-supervised learning method. Only a few tags are required to train a reliable model that decomposes the text into components on the target emotion component. The text generation part can generate text with specific emotional components based on the model annotated by the decomposition part. Compared with the current text style transfer work, the text generation model does not need to use neural codec, so it has higher efficiency and better interpretability.

The algorithm described in this paper is a set of forward-looking tool kernel, which has an interpretable reasoning process. It can be used for the tasks such as psychological counseling technology training and text generation of public opinion robot in the future.

Keywords: Natural language processing; Text generation; Multi-dimension text analysis; Language model; Semi-supervised learning

理论基础

心理空间与文本认知理论概述

认知语言学认为，语言是认知的表征，语言研究必须与认知研究结合起来。文本由语言构成，必然也涉及到认知过程，因此对于文本的研究也可以结合认知语言学的研究成果。1994 年，美国认知语言学家吉勒 •福科尼耶(Gilles Fauconnier)提出了心理空间理论，以人类的认知活动为基础探讨意义建构过程。按照福科尼耶的观点，意义建构涉及到人们在思考、行动或交流时进行的高级、复杂的心理活动。人们用语言进行相互交流的时候，他们通常会建立起由语言信息和当下语境触发的概念域，在这些概念域中进行意义的构建，从而实现理解。

虽然在庞大的语言体系中，涉及到的信息和原则不计其数，但这并不代表我们必须对所有的语言学规则与文化概念域进行建模才可以进行分析。认知语言学家认为，由于阅读涉及到对单字的辨识、对词汇的理解以及对句子的分析，在读者的注意力有限的情况下，人脑必须能够下意识地处理包括语言信息在内的各种视觉输入。而在大部分情况下，被下意识处理、且承载意义的最小语言元素为词汇。“承载意义的最小语素”意味着，表达需要通过词汇的组合进行实现，文本通过词汇建立起表层结构，以此来传达自己的内涵。而对于读者来说，基本语素与读者概念域中的相关概念联系到一起，读者就可以进行适当的意义建构（或称情绪感知）。由于基于本能的理解过程是下意识进行的，这种建构得到的并非是文本要表达的直接语义——这里涉及到本文所研究的问题与自然语言倾向性分析的不同点，认知语言学的观点认为，语言是隐喻化的，所以人们在理解表层语言的时候思维也是隐喻性的[4]。如句子“你总是嫌弃这个嫌弃那个的 总觉得所有人都喜欢你 换个头像就是爱你了”尽管它在逻辑上表达的是负面情感，但“嫌弃”、“喜欢”和“爱”却暗指一个强大的形象，强大的形象反而又与正面情感存在着概念链接——这是潜藏在直接语义之下的、基于概念域产生的、不受控的信息——不同于意图使机器分析结果接近人类判断的倾向性分析算法，本文研究的，基于本能的认知过程是反逻辑的，就像我们没有意识到大脑中的化学反应一样，我们也没有意识到自己在理解语言时所进行的意义建构。这种意义建构是通过下意识对人产生影响，但优秀的作家往往会透过词汇的选择性使用，利用这种下意识的反应传递自己想要表达的感情。

长久以来，文学家、语言学家通过使用理论方法解读一些诗歌或小说[5]，试图分析这种下意识反应，寻找更适合的写作方式或语言运用规律。心理学家与认知科学家则通过近红外脑成像[6]、眼动跟踪[7]等技术研究人在阅读过程中的生理

反应，试图寻找文本与人类认知间的桥梁。本文研究的方法通过自然语言处理技术自动化地建构文本意义，或者说识别文本对读者认知的影响。当我们需要从认知语言学与心理空间理论出发进行分析时，就需要确定读者所拥有的“概念域”。康德指出，文本中的概念如果不与读者的“前理解”联系起来就不会产生任何意义。基于概念域的理解就是指“读者对文本的理解涉及到他们对已获得知识相关信息的回忆和组织”。换句话说，在心理空间的理论框架下，任何文本都只是给读者提供一定的线索，使他们利用脑中的知识与常识来构建文本的意义。因此，概念域可以被类比为读者在文本解读过程中对脑海中背景知识空间的激活。这与目前自然语言处理研究中预训练语言模型所构建的语义空间不谋而合。因此本文所提出算法的思路即为，在输入文本上进行采样，结合语言模型提供的信息，将高维的输入分本分解为低维度的数个概念分量。这数个概念分量构成的低维度向量即为该概念域下读者对输入文本的直观意义建构，或者说输入文本对读者认知的影响。

值得一提的是，如果将模型输出视为文本对读者认知的施加的影响。那么基于目前对机器学习算法对抗性攻击的研究[8]，将该模型视为攻击的对象，我们可以构造合适的输入文本，使得它对读者产生我们想要的特定影响。在计算机运算速度的加持下，使用该算法进行这一工作的速度是文学家和语言学家所不能达到的。这将为写作辅助、自动舆情控制及一些心理学与认识科学研究提供重要的工具。

多维度情感分析相关工作分析

本文的方法将高维的输入分本分解为低维度的数个概念分量。在形式上，这与多维度和细粒度情感分析相似。由于多维度情感分析往往不能使用离散标注[9]。目前的标注一般基于情感空间理论[10]，如 Valence-Arousal-Dominance（VAD）模型，Evaluation-Potency-Activity（EPA）模型等。Valence 和 Evaluation 表示好坏，arousal 和 activity 表示人的唤起度，dominance 和 potency 表示控制力。Yeh 等[11]和 Yu 等[12]就构造了基于 VAD 模型的中文情感词典。然而此类情感空间模型是作为一个力求通用性的描述存在，并不一定能覆盖特定任务的情感标记需求。

细粒度情感分析又称为基于属性的情感分析。如“群里人才辈出 有的有钱有的有才 还有和我一样的小菜鸡”，从单纯情感倾向性分析的角度来看很难判断这句话的整体情感倾向，因为他赞扬了“群里的人才”，但吐槽了自己。目前对基于属性的情感分析已经取得了一些成果，但缺乏对隐式特征和隐式评论的挖掘。如“我也喜欢别人问我问题 但是基本都去问大神了 没我的份”这句话没有

表达"成绩不好"的相关字词，但是明显可以看出该句表达出了对自己成绩的负面情绪。细粒度情感分析中的隐式特征抽取针对这一问题进行研究。Li 等[13]抓取了数百万条微博数据构造了一个中文内隐情绪分析语料库，解决中国人情感表达含蓄，口语中缺乏明显情感词汇的问题。根据认识语言学研究，所有词语（基本语素）都会基于日常惯用法产生与其高频共现词的概念连接，因此不需要显式地进行隐式情绪分析。根据这一假设，Yang 等[14]的工作更贴近于本文的需求，他们提出了一种基于图-注意卷积神经网络的隐式情绪分析模型。使用图卷积神经网络来传播语义信息。注意机制被用来计算单词对情感表达的贡献。本文的算法参考了其思想，但对实现进行了简化，直接在预训练语言模型的语义空间传播语义标签（这样不需要依赖图神经网络的训练）。

文本风格迁移相关工作分析

文本风格迁移是近几年出现的新研究方向。其与同为文本转换的机器翻译问题的区别是，文本风格迁移缺乏平行语料数据进行训练。

直观的思路是分离文本的风格与内容，然后使用新的风格重新生成。主要结构采用生成对抗网络(GAN)与编解码器。编码器负责分离文本的内容表征和风格表征，带解码功能的生成器负责根据指定的风格与内容生成文本。鉴别器基于迁移强度、语义保留、结果流畅性等指标进行优化。其中，风格表征作为解码器输入的一般被称为"可控生成"，如 Hu 等[15]。而如果解码器只输入内容编码，那么针对不同风格需要训练多个解码器，如 Xu 等[16]的工作为多解码器。而 Fu 等[17]的工作两类模型均被包含在内。Dai 等[18]的工作不显式分离输入文本的风格，而是直接将其与另一个编码器训练出的风格编码共同输入 Transformer 进行生成，与鉴别器联合训练。

Dathathri 等[19]基于针对神经网络的白盒对抗性攻击的思路进行指定风格化的文本生成，其训练一个属性模型攻击预训练的文本生成模型，属性模型在生成的每个时间步计算对抗梯度值，并将其叠加到语言模型 RNN 的隐状态上。通过在优化项加入对抗梯度与语言模型隐状态的 KL 散度，提高生成结果流畅性。

Guu 等[20]的工作不显式分离风格标签，而是对编辑行为建模。其提出的"原型-编辑"模型训练过程为从语料库采样一个待改写的原型句子进行编码，然后引入一个编辑向量在神经网络（神经编辑器）中修改句子编码，以生成新的句子。优化过程控制每次编辑的语义改变量，并使相似的编辑向量执行相似的语义编辑，使得随着训练可以使得编辑向量捕获语义信息。

以上工作均使用了神经语言模型和编解码器结构，虽然可以对文本的语义进行表征，但生成表征和迁移文本的过程依然是不可解释的。此外，根据我们对这

些方法的实验分析。交叉对齐和多解码器往往会丢失源语句的内容。解码器倾向于生成一个频繁但仅与目标属性弱相关的句子。我们的结论是在保留原始内容和删除原始风格之间存在微妙的平衡，现有的编解码器方法往往会牺牲其中一个。

Xu 等[16]与 Li 等[21]的思路与本文算法反向过程改写的实际行为类似，其假设文本传递的信息体现在文本中的某些特定词汇上。通过改变这些词汇，就可能直接改变整个文本的属性。Xu 等[18]提出的模型分为多个模块，先采用预训练的情感分类器去除原句中的情感词得到中立句子，再使用 seq2seq 向中立句子添加情感词，其中不同情感使用不同的解码器。Li 等[19]的工作先使用统计方法在已标记的不同风格语料中统计表示风格的关键 n-gram。对于待修改的句子，首先删除其中的风格 n-gram，然后在目标情感语料中寻找与原句子最相近句子中的风格 n-gram，最后使用直接替换和拼接中立句子编码与目标风格 n-gram 编码后重解码两种方法进行生成。

半监督情感标记相关工作分析

在目前半监督自然语言处理的工作中，使用预训练语言模型[2,22]已成为引入域外数据的一种强大方式。如方澄等[23]提出了一种基于 BERT 词向量构建全局特征图的半监督文本情感分类方法。该方法使用 BERT 词向量构建全局异构图，基于词语共现关系在全局异构图上进行特征传递。最后将全局特征图输入到图卷积网络中使用已标记数据进行训练。但在很多时候下，计算也是一种稀缺资源，微调庞大预训练语言模型的代价可能是昂贵的。

也有人试图利用变分自编码器(VAE)进行自然语言处理任务上的半监督学习，主要是以 seq2seq 模型的形式[24,25]，这些模型使用针对序列的编解码器结构。虽然这些工作的实验结果良好，但仍有许多悬而未决的问题需要进一步研究：首先，鉴于已知 seq2seq VAE 的训练难度[26]，这种方法在实践中是否有用值得怀疑。其次，目前还不清楚在文本分类等任务中是否真的需要这种复杂的模型。

本文希望能够使用一种较为简洁和稳定的方法，在少量标记数据实现大规模语料的分解。根据以往的研究可以知道，神经网络形成的中间表示空间可以作为语义表征[27]，基于神经网络的语言模型也具备这种性质[28]。因此本文使用类似标签传播算法的方法，利用预训练语言模型的表示空间，但不进行任何新的神经网络训练，避免训练过程的昂贵代价和不稳定性。如张璞等基于标签传播算法的工作[29]先构建词和种子词的图，图上的边是基于词和词之间的统计信息获得。然后用标签传播的算法获得新词的情感信息。本文采用方法的不同点是直接在连续语义空间上进行标签传播，因此不需要构建图的过程。

正向过程：文本分解

概念定义

文本分解问题的定义为：将高维的输入分本分解为低维度的数个概念分量。这些概念分量通常描述的是该文本（在不同方向上的）情绪表达。首先介绍描述文本分解算法将用到的基本概念。

定义 1 基本语素

输入分解算法的文本信息被处理的最小单位。视不同问题可以被定义为“字”、“词”、“短语”等。一般认为，被下意识处理、且承载意义的最小语言元素为词汇[5]。本文的实验中也将基本语素定义为单词。

定义 2 基本语素的嵌入表示

语言形式的基本语素在语言模型中的对应数字化形式。我们熟悉的 word2vec 词向量是嵌入表示的一种。从各种语言模型中都可以提取对基本语素的嵌入表示（一般是网络中间层的计算结果），但不是所有语言模型都可以保证不同基本语素嵌入表示的正交性[30]，一般来说，捕获语义越丰富的语言模型单个嵌入间的正交性越差。需要根据分析需求选择语言模型。

定义 3 语义空间

描述各基本语素嵌入表示远近关系的空间。语义相似的基本语素将在语义空间上有着更近的距离。

定义 4 概念构成

文本分解算法的输出，是由数个**概念分量**（实数）构成的向量。在训练过程中，算法会计算每个基本语素的概念构成。在分解阶段，算法会根据输入文本中各基本语素的概念构成计算整个文本的概念构成。

标记-传播算法

“标记-传播”是文本分解算法的训练步骤，它是一种半监督方法，目标是让用户只需要进行少量的标记即可获得大部分基本语素相对准确的概念构成。由于在语义空间相近的基本语素具有相似的语义，因此应当具有相似的概念构成。所以，对于每个用户进行的基本语素标记，算法在语义空间向其近邻的基本语素传播本次标记信息，以实现对其它基本语素的自动标记。

具体来说，定义每个用户标记为三元组$Input$：

$$Input = (InputWord, InputCmptIndex, InputCmptValue)$$

其中$InputWord$为用户本次标记的基本语素，$InputCmptIndex$为被标记的概念分量类型，$InputCmptValue$为该概念分量被标记的值。

首先，算法会计算$InputCmptValue$相对目前$InputWord$该分量值的增量，如下式所示：

$$\Delta CmptValue = InputCmptValue - AllCmpt_{InputWord,\, InputCmptIndex}$$

然后将当前存储的基本语素该分量值设为目标值，如下式所示：

$$AllCmpt_{InputWord,\, InputCmptIndex} = InputCmptValue$$

其中$AllCmpt$为存储所有基本语素各分量值的矩阵。

由于每次传播，算法都会使得近邻的基本语素拥有相似的概念构成。因此在新的传播发生时，只需要向周围传播此次标记产生的增量$\Delta CmptValue$即可使得邻域的概念构成保持平衡。传播过程为一个递归过程，每次递归传播的增量发生衰减，直到小于阈值，传播停止。具体算法如下：

算法 1 “标记-传播”算法的递归传播过程

Propagate($word$,$CmptIndex$,$\Delta CmptValue$)

$NeighborhoodWordSet = FindNeighborhoodWord(word)$
For $word2$ in $NeighborhoodWordSet$
  If $word2$ not in $visitedWords$
    $dist = cosine(word, word2)$
    $\Delta CmptValue2 = dist \times \Delta CmptValue$
    $AllCmpt_{word2,\, CmptIndex} = AllCmpt_{word2,\, CmptIndex} + \Delta CmptValue2$
    If $\Delta CmptValue2 > \tau$
      $Propagate(word2, CmptIndex, \Delta CmptValue2)$

递归函数的参数为基本语素$word$，概念分量类型$CmptIndex$，概念分量增量$\Delta CmptValue$。$FindNeighborhoodWord$ 函数在语义空间中查找$word$的近邻基本语素，其实现不宜将“近邻”的阈值设置太小，因为算法的递归深度应由阈值τ控制。$cosine$函数计算余弦相似度，由于余弦相似度范围为$[-1,1]$，因此可以直接作为衰减权重。最终，当增量衰减到小于τ时，递归结束。

用于用户标注的交互式程序

本文设计了如图 1 所示的交互式命令行前端让用户更方便地进行标注（其中绿色部分为用户输入）。

```
loaded
word:query
word:语义
{'文艺': 0.06037968964049944, '严肃': 0.22748377307319434, '悲伤': 0.08485831520986183}
word:语义
dim:文艺
val:0.1
finished
```

图 1 用于标注的交互式程序

该交互式程序包含“标注”、“查询”与“保存”三个命令。交互式程序除了有助于用户准备训练数据外，更重要的一点是可以让用户修正标注信息传播时产生的一些错误。假设在语义空间上正确的概念构成分布应为图 2 的降维可视化结果所示（颜色代表语义强度）。

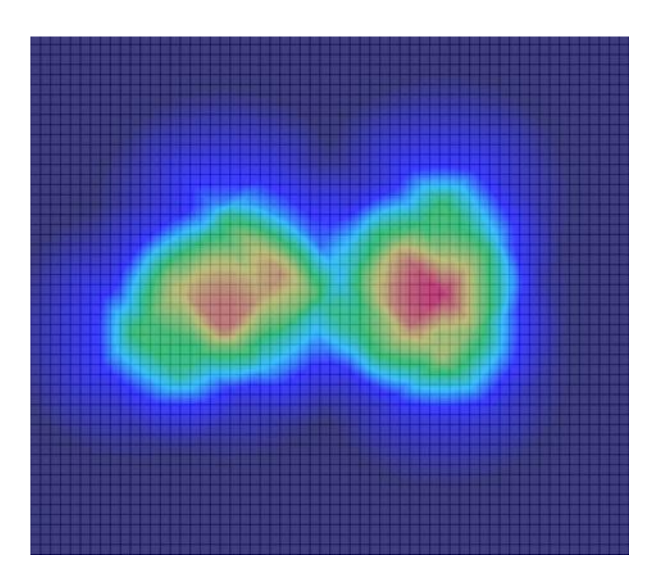


图 2 真实概念构成分布

但如果用户的标注往往是不完善的，如果使用如图 3 的标注数据，可能会导致传播过程过度强化了左侧区域。

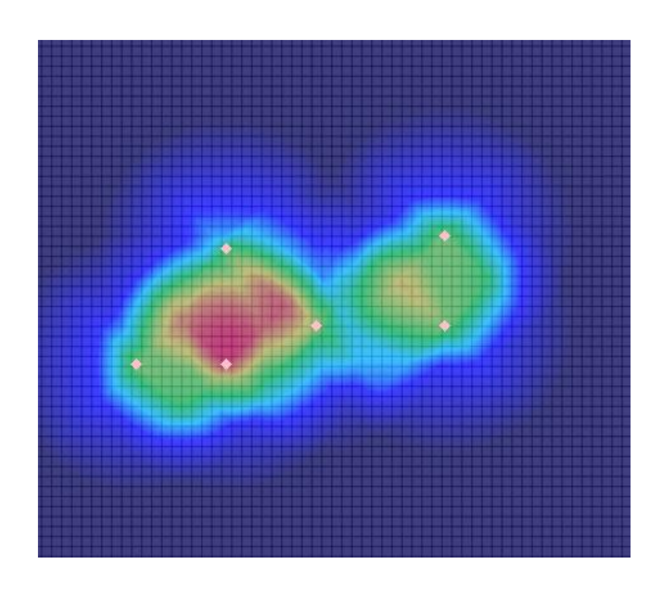


图 3 不完整的用户标注导致传播失真

此时用户可以使用交互式程序查询到右侧区域与左侧区域的概念构成不符合常理，然后选择补充标注右侧区域或修改左侧区域的标注。由于“标记-传播”算法在语义空间中传播的是增量，那么当修改被过度强化的左侧区域时，算法会向周围传播负增量，得到正确结果。如图 4 所示。

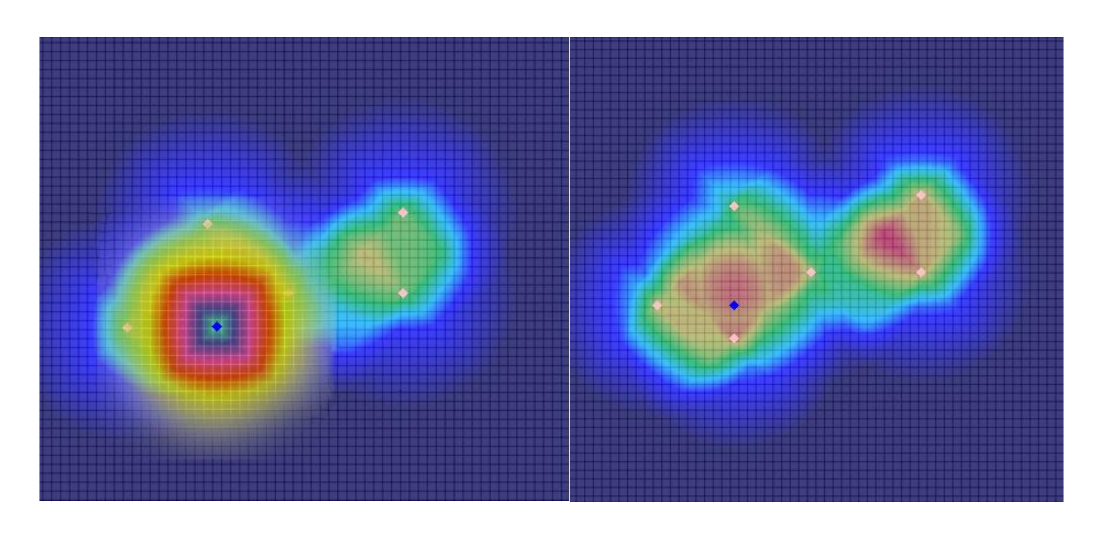


图 4 通过修改左侧区域标记修正失真

其中左侧图为重新标记蓝色菱形位置时产生的负增量传播范围（负增量数值大小用反相颜色表示，即蓝色菱形周围的红黄色区域），叠加负增量后的语义空间概念构成分布为右图所示。可以看出，它更接近于真实概念构成分布。

这样，当用户发现算法结果存在问题时，可以直接使用交互式程序查询可能出现问题的基本语素，并快速进行针对性修正。而不像大部分机器学习模型只能通过机械性地添加数据进行完善。

文本分解算法

完成对语义空间的标记后，即可对真实语料进行文本分解。最直接的方法是计算语料中所有基本语素的概念构成均值作为该语料分解结果。首先使用分词算法拆分输入分本后，再使用停用词表去除停用词。对于剩余的每个单词$TextWord_i$，查询其对应的基本语素嵌入表示$word_i$，如果可以直接查到，那么可以直接用于计算均值，如下式所示：

$$AvgVec = \Big(\sum_{word_i \in allWord} word_i\Big)/n$$

其中 n 为本次处理的单词数量。

对于无法直接查询到嵌入表示的$TextWord_i$（未登录词），本文采用类似FastText[31]的嵌入合成方法。如对于中文单词，$TextWord_i$为字序列$\{c_1, c_2 ...... c_n\}$。对于其中的每个字c_i，可以对其出现的所有可查询到的词嵌入进行平均，作为该字的嵌入表示，如下式所示：

$$EmbeddedC_i = \Big(\sum_{c_i \in TextWord_j} word_j\Big)/n$$

得到$\{c_1, c_2 ...... c_n\}$中所有字的嵌入表示后，可使用其均值作为$TextWord_i$的嵌入表示$word_i$，如式下式所示：

$$word_i = \Big(\sum_{c_i \in TextWord_i} EmbeddedC_i\Big)/n$$

由于使用 FastText 方法计算的词嵌入并非经过语言模型严格训练得到，因此需要降低其在文本概念构成均值计算中的权重，本文实验中将其权重设为 0.5。整体算法如下：

算法 2 直接均值法文本分解

输入： 单词序列$TextWord$

输出： 概念构成$AvgVec$

1. $n = 0, AvgVec = \{0,0, ...0\}$
2. For $TextWord_i$ in $TextWord$
3.   If $word_i$ (Embedded representation of $TextWord_i$) cannot be found
4.   Calculate $word_i$ using FastText
5. $n = n + 0.5$
6. $AvgVec = AvgVec + word_i \times 0.5$

7. Else
8. $n = n + 1$
9. $AvgVec = AvgVec + word_i$
10. $AvgVec = AvgVec/n$

可以看出，该方法并没有考虑基本语素间的顺序关系。这是因为，根据认知科学研究，人在阅读时进行的是跳读，而并非从左到右的机械扫描。这意味着，一个短句中的所有词语会并发地被下意识接收，然后再基于逻辑解析它的顺序关系。所以，基于本文的研究目的，不需要考虑短句中的顺序关系，每个基本语素应视为具有相同的地位。而对于长文章，则需要将其拆分成短句并对每个短句进行分析，这样可以得到整个文章情绪表达变化的动态过程。需要注意的是，尽管人在阅读短句时是并发的，但在阅读长文章时是顺序的，因此每个短句的情绪表达会受其前驱短句的影响。设长文章为短句序列$\{s_1, s_2 \ldots\ldots s_n\}$组成，对于$s_i$的概念构成$EmbeddedS_i$，如下式所示：

$$EmbeddedS_i = r \times EmbeddedS_{i-1} + (1 - r)RawEmbeddedS_i$$

其中$RawEmbeddedS_i$是将s_i输入到算法 2 得到的结果。r 为折扣因子。

实验

由于文本分解算法得到的是由文化产生的“概念域”所建构的深层语义，因此本章的实验将针对文化现象进行分析。

在当前的恋爱文化中，一方想要发展关系的程度往往与另一方外貌紧密相连。如我们曾经使用“当你在一对一聊天中夸奖对方可爱时，你是否并不排斥与对方发展关系？”这一问题进行问卷调查，最终平均有八成的受试者选择了“是”，其中女性受试者超过九成都选择了“是”。这意味着，在一对一交流中，“可爱”、“真好看”、“好帅”等词往往兼顾了“外貌”和“发展关系意愿”两个概念分量上的语义强度。而与“外貌”相对的则是“内在”。那么是否存在类似的表达，可以兼顾“发展关系意愿”和“对对方内在的赞美”呢？

本节的实验首先使用文本分解模型分析这一问题，然后与人类受试者提出的方案进行对比。

基于文本分解的分析

首先介绍实验设置。本实验设定的概念构成含有“外貌评价”、“发展关系意愿”、“内在评价”三个分量；预训练模型来自使用微信公众号文章数据训练的 word2vec 模型。训练数据来自网络收集的表白、表白试探语料和知乎问题“你

见过最聪明的男孩子/女孩子是什么样子”下的回答两部分。由人工对整个句子的三个概念分量进行标注，随后使用摘要算法抽取每个句子中的1-3个关键词作为实际被标记的基本语素，使用“标记-传播”算法进行半监督训练，共进行50次标记。部分标注数据如表1所示。

表1 部分标注数据

语料	**标注**（[外貌,发展关系意愿,内在]）
你果然嫌弃我，我刚才只是在试探	[0,5,0]
特别是对于在意的人就会觉得很委屈	[0,5,0]
就是觉得你好可爱	[8,5,0]
你怎么还不跟我表白	[0,9,0]
知识真的可以让一个人很有魅力	[0,3,5]

其中蓝色部分为摘要得到的关键词，它们作为被标记的基本语素。

算法使用Python3.6实现。为了模拟低计算资源环境，训练在AMD Athlon2 X2 640处理器、8G内存的PC平台上进行。传播停止阈值$\tau = 0.1$。所用训练时间为1h 2min 38sec。

首先进行样本内精度测试。例句“就是觉得你好可爱”，归一化（使概念构成模为10）分解结果为*[2.071,2.386,0.125]*；例句“知识真的可以让一个人很有魅力”分解结果为*[1.722,0.346,2.789]*（三个分量顺序为*[发展关系意愿,外貌,内在]*），各分量比例符合预期。

本节的目标是使用文本分解模型试图寻找类似“可爱”（概念构成为*[7.0,8.172,0.433]*），兼顾“发展关系意愿”和“内在评价”的最佳表达。方法是设定起点词，以其嵌入表示坐标为起点，以近邻的50个点估算区间梯度，使用贪心法向提升目标分量方向移动，寻找合适的表达。

使用“可爱”为起点词，提升“内在评价”分量，结果如图5所示。

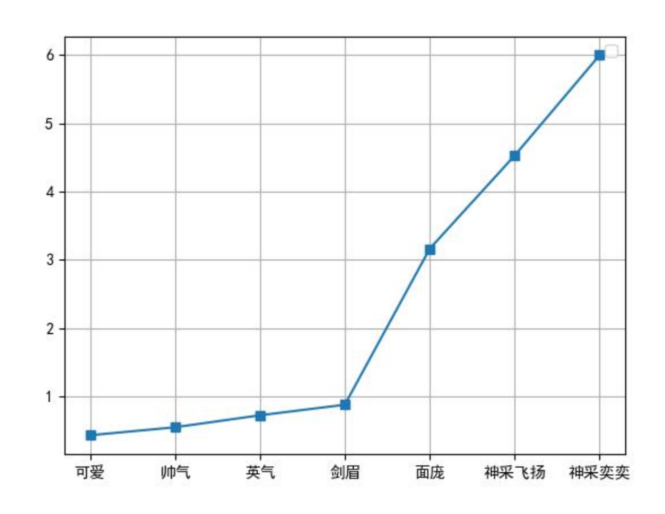


图 5 以“可爱”作为起点词迭代到局部最优

其中横坐标为每次迭代产生的更适表达，纵坐标为“内在”分量的语义强度。可以看出，虽然随着迭代的进行，词语的语义向“开朗乐观”方向移动，但仍在借助外貌特征进行表达。

使用“魅力”为起点词，提升“内在评价”分量，结果如图 3-5 所示。

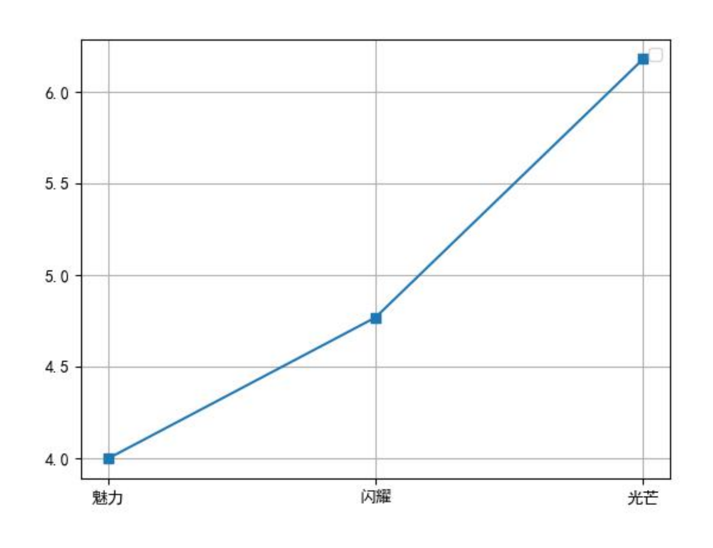


图 6 以“魅力”作为起点词迭代到局部最优

在我们的文化中，描述一个人“闪光”通常含有对对方能力赞美和崇拜的含义，而“崇拜”具有“发展关系意愿”分量上的语义强度，这是一个可以接受的结果。

使用“知识”为起点词，提升“发展关系意愿”分量，结果如图 3-6 所示。

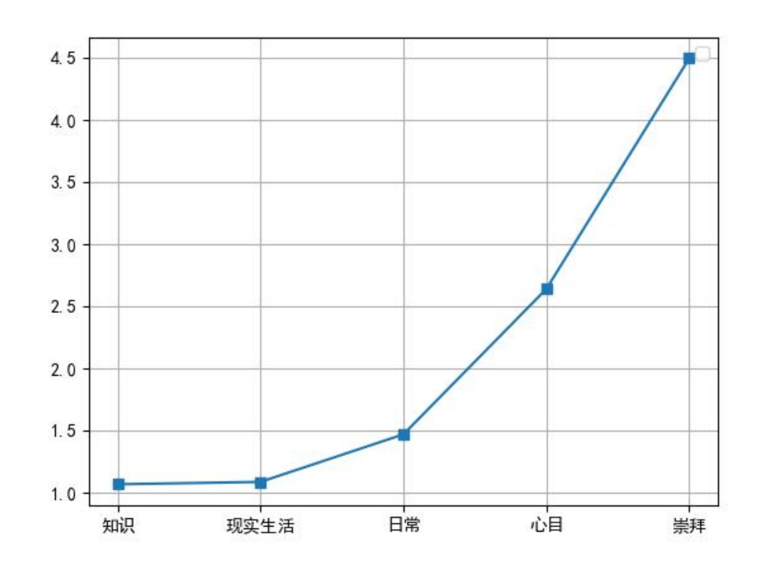


图 7 以“知识”作为起点词迭代到局部最优

本次迭代收敛到“崇拜”和上一次收敛到“闪光”的结果可以互相印证，因为这两个词具有较高的共现频率，明显可以形成简单句“很崇拜（你这种）闪闪发光的人”。说明本章提出的算法所标记的语义空间与人类认知相符，且达到了本节的任务目标。

人类受试者方案分析

在上一节，文本分解算法的分析结果显示“崇拜”、“闪光”可以兼顾“发展关系意愿”和“对对方内在的赞美”两个分量上的语义强度。本节将收集人类的结果进行对比。我们在在线平台以“用尽量简短的句子表达因对方内在美而生的喜欢”为问题收集答案。对得到的结果使用文本分解模型和读者主观评测两种方法进行评价，部分结果如表 2 所示。

表 2 部分人类受试者方案评价

答案语料	**文本分解结果**（[发展关系意愿,外貌,内在]）	**读者主观评测**
跟你相处真的好快乐，我感觉之前从来没遇到像你一样让我聊天舒畅开心过的人	[2.322,0.254,0.152]	聊天开心是两个人之间的，和对方的内在没有必然联系吧
跟你这么可爱的人聊天真的好开心	[2.358,2.227,0.174]	没体现出内在
跟你这么懂我的人聊天真的	[2.343,0.271,0.176]	“懂我”不一定代表内在美

很开心		吧……
我感觉你对我来说特别可爱，我很喜欢你这种性格	[2.598,2.472,0.417]	太绕了
我感觉我们简直是天造地设的一对	[2.364,0.434,0.093]	没体现出内在
也许你就是我人生拼图中另一半	[2.632,0.535,0.487]	没体现出内在

对比表中的文本分解结果与主观评测结果，除了句子“我感觉你对我来说特别可爱，我很喜欢你这种性格”算法无法理解句中的“可爱”是修饰“性格”，因此导致算法结果外貌分量较高，不符句子本意外。其它句子的分解结果均与主观评测结果相符，证明了算法可以有效分析句子对读者的认知的影响。用同样的方法评价算法给出的结果，如表 3 所示。

表 3 算法方案评价

答案语料	**文本分解结果**（[发展关系意愿,外貌,内在]）	**读者主观评测**
很崇拜你这种闪闪发光的人	[2.232,0.049,2.102]	相对来说好一些，但“闪光”指的是能力，能力只是内在的一部分

可以看出，本节问题下，算法方案的两种评价结果均优于人类方案。可以证明，算法结果构造的简单句更贴近我们需要的表达。

反向过程：文本生成

文本生成算法

基于上一章的工作，我们已经可以使用文本分解算法分析文本的情绪表达。那么既然已经知道语义空间中各个基本语素的概念构成，是否可以让算法自动撰写（修改）文本，表达我们想要传递的特定情绪或暗示？这一问题程序化地表述即为：如何以优化文本分解算法的输出为目标，调整算法输入。这即是本章“反向过程”的含义。

本文设计的文本生成算法基于一个重要的结论：文本属性的转移通常可以在保持句子的其余部分基本不变的情况下，通过改变句子中表示特定属性的几个语素来完成。以往的认知科学[7]和自然语言处理[21]研究都证明了这一点。

在针对机器学习模型的对抗性攻击研究中，一种常见的思路是，计算模型损失，估计模型内部在当前输入下的梯度，并将输入向梯度上升方向移动[8]。对于本章的问题，由于已知整个语义空间的概念构成分布，因此可以计算每个位置的各分量梯度。那么可以以待改写文本中的基本语素为起始点，根据梯度方向，向目标概念构成移动，在新的区域选择更符合目标的基本语素。

形式化地表述，设语义空间为SS，此时我们可以查询嵌入坐标e的概念构成：

$$SS(e) = W$$

其中概念构成$W = \{w_1, w_2 ...... w_n\}$。由于概念构成的每个分量都是正交的，因此可以将查询多个分量的过程视为多个单因变量函数。那么当需要提升（或降低）分量w_i时，就可以向函数$SS_i(e) = w_i$的梯度（或负梯度）方向移动。即对于待改写句子中的每个基本语素嵌入$word_i$，当想要增大其第j个概念分量到值v时，迭代公式如下式：

$$word_i^{t+1} = word_i^t + \alpha \nabla SS_j(word_i^t)$$

其中$word_i$的上标为迭代次数，α为步长。迭代停止条件如下式：

$$|SS_j(word_i^t) - v| < \tau$$

在该嵌入坐标附近采样一个基本语素$TextWord_i$作为原先位置的替换。

当需要同时优化多个概念分量时，每步的前进方向需要叠加各分量的梯度。如下式所示：

$$word_i^{t+1} = word_i^t + \sum_j \alpha \nabla SS_j(word_i^t)$$

此时停止条件计算所有分量距离目标差的平均，如式下式所示：

$$\left(\sum_j |SS_j(word_i^t) - v_j|\right) / n < \tau$$

这种直接替换方法很多时候会产生不恰当的上下文。本文采用两种方法对其进行约束。首先限制替换的基本语素词性与原先相同（保证语法正确性）；其次记录迭代中每一步选择的替换结果，对每个位置、每一步结果构成的所有生成文本组合在语言模型中计算联合概率，选择概率更高的结果（保证语言流畅性）。这里比较适合使用的语言模型是刘廷镇等[32]提出的概率图语言模型，因为该模型可以直接计算指定句子出现的概率，而神经网络语言模型的联合概率信息大多是隐藏的。

实验

本节的实验使用文本风格迁移任务来检验本章算法的性能，我们尝试将小说文本迁移到相反的风格。首先介绍实验设置。本实验设定的概念构成含有“悲伤”、“严肃”、“文艺”三个分量；预训练模型来自使用微信公众号文章数据训练的 word2vec 模型。训练数据来自对 lofter 平台随机抓取的小说文本进行摘要得到的关键词，并对关键词的概念构成进行人工标注，使用“标记-传播”算法进行半监督训练，共进行 51 次标记。

测试 1：改写轻小说

本节针对口吻轻松的轻小说文本，因此提升其“严肃”分量。共实验五个段落，前段为原文本，后段为改写后的文本，括号部分为改写处。

段落 1：

“李腾，你要是再跑，我就开枪了!”后面一个戴着眼镜的黑衣人一边喘着粗气一边朝着前面那个男孩喊道。正在前面疯跑的李腾听到这话直接不屑地“呸”了出来，心想傻子才束手就擒停下来任你们宰割。这一群疯子找他李腾也不知道是怎么回事，如今已经 2021 年，居然还有这么一群人存在，真是匪夷所思。

“李腾，你要是再跑，我就【枪杀你】！”后面一个佩戴着眼镜的黑衣人一边【泻着】粗气一边朝着前面那个【孩子喝道】。正在前面【发疯跑】的李腾听到这话直接嘲讽地“呸”了出来，【心道】傻子才束手就擒【停住】任你们【残杀】。【这群政客】找他李腾也不知道是怎么回事，如今已经 2021 年，居然还有这么一群人存在，【当真是毛骨悚然】。

“开枪”改写为“枪杀”、“喘着”改写为“泻着”、“喊道”改写为“喝道”、“心想”改写为“心道”、“宰割”改为“残杀”、“真是匪夷所思”改写为“当真是毛骨悚然”，均是用书面化表达替代口语表达。由于严肃的表达往往是书面化的，因此书面化表达具有和“严肃”间的概念链接，这些改写都比较合理。“疯子”替换为“政客”发生了语义变化，是算法没有将语义保留作为约束所致。

段落 2：

李正存早先不学无术为非作歹，后来遇上了自己的师公李清源方被教化皈依道教，此后正德修念，一心清净向学

李正存早先不学无术为非作歹，后来遇到了自己的师公【薛清源】方被【教诫受戒佛教】，此后正德修念，一心【具足】向学

“李清源”改写为“薛清源”暗示着不同姓氏在认知上的微妙差异。这也是为什么玄幻小说角色多使用“顾”、“林”、“萧”等姓氏而非常用姓。“佛教”改写为“道教”可能是因“佛教”处在热点标记区域附近，“道教”较远，因此各个概念分量强度都较弱。“清净向学”改写为“具足向学”可以被视为一种防止常用语羞涩的处理（常用语可能在“严肃”分量上被传播了较多的负增量）。

段落 3:

黎祈不紧不慢的跟着，手中夹着一本《谐音学日语速成秘籍》。他虽然重新变成了肉体凡胎，但是精神强度并没有改变，所以学起东西快的吓人，只要默念一遍就能记下。他甚至在想，如果阿福把他送到高考前一周或许更好，他没准能顺道破个总分记录

黎祈不紧不慢的跟着，手中夹着【一册】《谐音学英语速成秘籍》。他虽然重新变回了肉体凡胎，但是【神性稳定性】并没有【转变】，所以学起东西快的吓人，只要默念一遍就能记下。他甚至在想，如果阿福把他【送来】高考前一个月或许【更佳】，他没准能顺道破个总分记录

“一本”改写为“一册”、“更好”改写为“更佳”是用书面化表达替代口语表达。“精神强度”改为“神性稳定性”从感知上提升了“严肃”分量。

段落 4:

上个时空他没有参加学校组织的赴日游学，和开了狂暴的吕布似的和各种嗜血怪物生死相搏。

上个【幻境】他没有参与【学生】组织的赴日【亲子游】，和开了狂暴的【董卓】似的和各种嗜血【吸血鬼寂灭相搏】。

“生死相搏”改写为“寂灭相搏”可以被视为一种防止常用语羞涩的处理。其它改写均没有可被感知的依据。

段落 5:

禁忌品是一系列受群星污染而产生诡异效果的物品总称，凡是接触过它的生物都会间接遭到污染，幸运者觉醒异能，不幸者失控变成怪物

【忌品】是【多方面】遭受群星污染而产生【可怖】效果的【产物】总称，凡是接触过它的生物体都会间接遭到污染，幸运者【觉知恶魔】，不幸者失控变成【吸血鬼】

“一系列”改写为“多方面”、“物品”改写为“产物”、“生物”改写为“生物体”均为换用更专业的术语提升“严肃”分量。“诡异”改写为“可怖”为换用书面表达。“觉醒异能”改写为“觉知恶魔”可以被视为防止常用语羞涩的处理，但使得句子发生了语义变化。

改写心理小说

本节针对心理小说文本。由于此类题材一般进行较为沉重的自我剖析，因此我们反其道而行则要降低其“悲伤”分量。共实验八个段落，前段为原文本，后段为改写后的文本，括号部分为改写处。

段落 1：

我开始真正思考一些遥远而抽象的问题。例如，离开这座小镇要怎么乘车？如果一架空中的飞机同时处于两个国家，一个国家是秋天，另一个是冬天，那么飞机上是什么季节？

我开始真正【看待】一些【实际】又【具象】的问题。例如，【回到】这座小镇要怎么【坐火车】？【要是】一架空中的【客机】同时处于【俩】国家，一个国家是秋天，另一个是冬天，那么【客机】上是什么【时候】？

这些改写均为直接使用反义词，因为这些词与原先词之间具有负相关性，因此在传播阶段直接被传递了负增量，符合使“悲伤”分量降低的目标。“乘车”改写为“坐火车”、“两个”改写为“俩”、“季节”改写为“时候”均为使用更口语化的表达。因为较强的书面化表达往往与“感怀伤时”有着相关性，反之，口语化表达则可以降低“悲伤”分量。因此这些改写均比较合理。

段落 2：

那个十几岁的我，对于这种离别，理解得过于浅薄，思考得又过于深刻。

那个十几岁的我，对于这种离别，【接纳】得【特别庸俗】，【看待】得又【颇深】。

“过于浅薄”改写为“特别庸俗”、“思考”改写为“看待”、“过于深刻”改写为“颇深”均为使用书面化程度更弱的表达。

段落 3：

当被一个粉笔头砸醒时，那个瞬间，一个季节突然就消失不见。我睁开眼，再去怎么想，都只是听见一片笑声。

当被一个粉笔头砸【入睡】时，那个瞬间，一个【时节】突然就【没了】。我【睁眼】，再去怎么【打算】，都只是听到一片【尖叫声】。

直接将“醒”使用反义词改写为“入睡”造成了逻辑语病，通过加强检查联合概率的语言模型可以一定程度减少此类错误。“消失不见”改写为“没了”、“睁开眼”改写为“睁眼”、“想”改写为“打算”均为使用更口语化的表达。

段落 4：

早上七点，天还稍凉时我们就已经出发，我们像是进入了幻境，一个身着浅蓝衣服的尼姑正在寺庙外清扫空地。

【中午六点钟】，天还【比较】凉时我们就已经【到达】，我们像是进入了【幻化】，一个身着【深蓝】衣服的尼姑正在【佛像】之外清扫【菜地】。

“早上”被改写为“中午”涉及到一些文化知识：“凌晨”、“傍晚”等具备过渡界限性[33]的时间点常常会因带有对逝去的不舍和对未知的不确定性而带上负面情感。而有着太阳的“中午”则因不具备过渡界限性使得情感较为中立。因此此处改写是符合目标的。但后面无论接“七点”还是“六点钟”都会造成逻辑语病。“出发”改写为“到达”同样是因为相比“出发”与“离别”之间的概念连接，“到达”的情感更为中立。“稍”改写为“比较”使用了书面化程度更弱的表达，“空地”改写为“菜地”则使用了更为市井化的意象，二者都有助于降低“悲伤”分量。

段落 5:

有一天躺在凉席上，窗户栏杆上的贝壳风铃被吹响

有一天【斜靠】在凉席上，【阳台雨棚】上的贝壳风铃被吹响

“窗户栏杆”改写为“阳台雨棚”使用了更为市井化的意象，有助于降低“悲伤”分量。

段落 6:

我学会了吸烟。契机是大二时交往过的一个学长，他说其实我们俩也只是在某一个路口，只是在某一个路口相遇，没有相同的过去，未来也渺茫

我学会了【抽烟】。【切入点】是大二时【合得来】过的一个【同学】，他【开玩笑】其实我们俩也只是在某一个【街口】，只是在某一个【街口相爱】，没有【一样】的过去，【今后】也渺茫

“吸烟”改写为“抽烟”、“交往”改为“合得来”、“相同”改写为“一样”均使用了更口语化的表达。“契机”改写为“切入点”、“路口”改为“街口”、“未来”改为“今后”均使用了文学化程度更弱的表达。“说”改写为“开玩笑”直接使用“悲伤”分量为负的表达。

段落 7:

我记得初中的一个下午，我逃课回家经过这里时，亲眼目睹过一起车祸现场

我记得初中的一个【周六】，我【自习】回家经过这里时，亲眼目睹过一起【事故】现场

“逃课”改写为“自习”使用了相对更中立的意象。“车祸”改写为“事故”减轻了语义强度。“周六”作为休息日相比其它时间点“悲伤”分量更低。因此这些改写均比较合理。

段落 8:

在那一百多秒里，我发现，原来什么都没变，无论是向外看，还是向内看，我的全世界都停止在这短暂又漫长的一百多秒中，而当绿灯亮起时，能够改变的却只有我

在那一百多秒【里头】，我发现，原来什么都没变，【还】是向外看，还是向内看，我的全世界都【中止】在这【欢愉】又漫长的一百多秒中，而当绿灯亮起时，能够【决定】的却只有我

“里”改写为“里头”使用了更口语化的表达。“无论”改写为“还”使用了文学化程度更弱的表达。“短暂”改写为“欢愉”直接使用“悲伤”分量为负的表达。“改变”存在着与困难的概念链接，相比之下“决定”是更中立的表达。

对以上两部分实验改写处的逐个分析计数可以得到，正确改写、错误改写和无可感知意义的改写比例为 48:3:19，可以证明算法的有效性。

结论

本文提出了一种文本分解算法用于分析文本的深层语义，填补了目前自然语言处理研究中对深层语义、下意识认知等问题研究的不足。该方法不需要微调复杂的神经网络语言模型，而是在预训练语言模型所构成的语义空间上使用传统的半监督学习方法。这使得算法既可以利用神经语言模型的数据优势，又可以稳定地实现小样本学习。由于其仅需要很少的标注数据，因此可以允许用户自定义不同的分解目标，可以覆盖很多不同领域的文本表达分析实际问题。

对训练出的文本分解模型反向使用该算法，可以实现改写文本，使得其表达、强化或弱化特定的情绪分量（概念分量）。相比目前的文本风格迁移研究，本文的算法具有相当低的计算代价，但却可以得到更加丰富的结果。

目前实验部分仅使用了较为简单的 word2vec 语言模型构建语义空间。如果今后使用嵌入表示与上下文相关的 BERT、GPT-3 等语言模型，构建语义空间及空间梯度的计算代价将会是值得关注的问题。如果想要继续保持方法的效率优势，可能需要研究一些相关数据结构与重复结果利用的优化方法。

本文的方法可以作为舆情机器人、写作辅助程序、心理咨询辅助程序等大量具有实际价值工具的内核。且本文方法的思路并不只局限于自然语言处理领域，可以向其它模态扩展，如尝试使用目标检测预训练模型分析观察者对图像信息的下意识感知。

访问实验程序

sg-first/Language-Decomposition: Natural Language Decomposition Algorithm for Detecting Deep Semantics (github.com)

参考文献

[1] Sapach S. Tagging My Tears and Fears: Text-Mining the Autoethnography[J]. Digital Studies/le Champ Numérique, 2020, 10(1), DOI:10.16995/dscn.328.

[2] Brown T B, Mann B, Ryder N, et al. Language Models are Few-Shot Learners[C] // Proceedings of 2020 Neural Information Processing Systems. Cambridge: Harvard Press, 2020, arXiv:2005.14165.

[3] Liu T, Zhou T, Shi Y, et al. A Herd Effect Detection Method Based on Text Features[C] // Proceedings of 2022 International Conference on Natural Language Processing. Xian: IEEE, 2022, DOI:10.20944/preprints202111.0499.v1.

[4] 赵霞. 概念合成:语言隐喻化的认知过程[J]. 华东船舶工业学院学报：社会科学版, 2004, 4(3):60-64.

[5] 王宣又. 心理空间与文学文本意义的认知探寻[D]. 重庆:四川外语学院, 2011.

[6] Li C, Ding K, Zhang M, et al. Effect of Picture-Book Reading With Additive Audio on Bilingual Preschoolers' Prefrontal Activation: A Naturalistic Functional Near-Infrared Spectroscopy Study[J]. Frontiers in Psychology, 2020, DOI:10.3389/fpsyg.2020.01939.

[7] Cyw A, Mjta B, Ccta B . Predicting Cognitive Structures and Information Processing Modes by Eye-tracking when Reading Controversial Reports about Socio-scientific Issues[J]. Computers in Human Behavior, 2020, 112(1): 112-118.

[8] HanXu, YaoMa, Hao-ChenLiu, et al. Adversarial Attacks and Defenses in Images, Graphs and Text: A Review[J]. International Journal of automation and Computing, 2020(2):151-178.

[9] Li M, Qin L, Long Y, et al. Inferring Affective Meanings of Words from Word Embedding[J]. IEEE Transactions on Affective Computing, 2017, 8(4): 443-456.

[10] Verma G K, Tiwary U S . Affect Representation and Recognition in 3D Continuous Valence-arousal-dominance Space[J]. Multimedia Tools and Applications, 2017, 76:2159-2183.

[11] Yeh J F, Kuang T Y, Huang Y J, et al. Dimensional Sentiment Analysis in Valence-arousal for Chinese Words by Linear Regression[C] // Proceedings of 2016 International Conference on Asian Language Processing. Tainan: ACM, 2016, DOI:10.1109/IALP.2016.7875998.

[12] Yu L C, Lee L H, Hao S, et al. Building Chinese Affective Resources in Valence-Arousal Dimensions[C] // Proceedings of 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego: ACL, 2016, 540-545.

[13] Li D, Wang J, Zhang X . CIEA: A Corpus for Chinese Implicit Emotion Analysis[C] // Proceedings of 2019 International Conference on Asian Language Processing. Shanghai: COLIPS, 2019, DOI:10.1109/IALP48816.2019.9037667.

[14] Yang S, Xing L, Li Y, et al. Implicit Sentiment Analysis based on Graph Attention Neural Network[J]. Engineering Reports, DOI:10.1002/eng2.12452.

[15] Hu Z, Yang Z, Liang X, et al. Toward Controlled Generation of Text[C] // Proceedings of 2017 International Conference on Machine Learning. Sydney: IMLS, 2017, arXiv:1703.00955.

[16] Xu J, Sun X, Zeng Q, et al. Unpaired Sentiment-to-Sentiment Translation: A Cycled Reinforcement Learning Approach[C] // Proceedings of 2018 Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018, arXiv:1805.05181.

[17] Fu Z, Tan X, Peng N, et al. Style Transfer in Text: Exploration and Evaluation[C] // Proceedings of 32th AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018, arXiv:1711.06861.

[18] Dai N, Liang J, Qiu X, et al. Style Transformer: Unpaired Text Style Transfer without Disentangled Latent Representation[C] // Proceedings of 2019 Annual Meeting of the Association for Computational Linguistics. Hawaii: ACL, 2019, arXiv:1905.05621.

[19] Dathathri S, Madotto A, Lan J, et al. Plug and Play Language Models: A Simple Approach to Controlled Text Generation[J]. arXiv preprint, 2019, arXiv:1912.02164.

[20] Guu K, Hashimoto T B, Oren Y, et al. Generating Sentences by Editing Prototypes[J]. arXiv preprint, 2017, arXiv:1709.08878.

[21] Li J, Jia R, He H, et al. Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer[C] // Proceedings of 2018 Conference of the North American Chapter of the Association for Computational. New Orleans: ACL, 2018, arXiv:1804.06437.

[22] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. arXiv preprint, 2018, arXiv:1810.04805.

[23] 方澄, 李贝, 韩萍. 基于全局特征图的半监督微博文本情感分类[J]. 信号处理, 2021, 37(6):1066-1075.

[24] Yang Z, Hu Z, Salakhutdinov R, et al. Improved Variational Autoencoders for Text Modeling using Dilated Convolutions[C] // Proceedings of 2017 International Conference on Machine Learning (ICML'17). Sydney: ACM, 2017, arXiv:1702.08139

[25] Xu W, Sun H, Deng C, et al. Variational Autoencoders for Semi-supervised Text Classification[C] // Proceedings of 30th AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016, arXiv:1603.02514.

[26] Bowman S R, Vilnis L, Vinyals O, et al. Generating Sentences from a Continuous Space[C] // Proceedings of 20th Conference on Computational Natural Language Learning. Berlin: ACM, 2016, arXiv:1511.06349.

[27] Radford A, Metz L, Chintala S . Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[C] // Proceedings of 2016 International Conference on Learning Representations. San Juan: ICLR, 2016, arXiv:1511.06434.

[28] Prusa J D, Khoshgoftaar T M . Improving Deep Neural Network Design with New Text Data Representations[J]. Journal of Big Data, 2017, DOI:10.1186/s40537-017-0065-8.

[29] 张璞, 王俊霞, 王英豪. 基于标签传播的情感词典构建方法[J]. 计算机工程, 2018, 44(5):6.

[30] Wang Q, Liu P, Zhu Z, et al. A Text Abstraction Summary Model Based on BERT Word Embedding and Reinforcement Learning[J]. Applied Sciences, 2019, 9(21):4701.

[31] Athiwaratkun B, Wilson A G, Anandkumar A . Probabilistic FastText for Multi-Sense Word Embeddings[C] // Proceedings of 2018 Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018.

[32] 刘廷镇, 张华. 基于变种概率图模型的文本生成算法[J]. 计算机应用, 2018, 38(A01):5.

[33] 田颖. 恐惧之源——《心是孤独的猎手》的"阈限空间"阐释[J]. 国外文学, 2015(2):10.